34.3 A 72Mb Separate-I/O Synchronous SRAM Chip with 504Gb/s Data Bandwidth

Chih Tseng[1], Jae-Hyeong Kim[1], Suzanne Chen[1], Mu-Hsiang Huang[1], Chungji Lu[1], Ikuo Hashiguchi[1], Yoshifumi Miyazima[2], Masahiro Ichihashi[2], Kentaro Maki[2], Katsuya Nakashima[2], Patrick Chuang[1].

[1]Sony Electronics, San Jose, CA
[2]Sony Semiconductor, Nagasaki, Japan

A 72Mb 144 separate-I/O CMOS SRAM operates at 875MHz DDR with 504Gb/s total BW. The 196mm^2 chip is fabricated in a 90nm 6M CMOS process. Dual read/write (R/W) self-timed clocks with core emulators are multiplexed to operate the SRAM core at 875MHz with random R/W in parallel per cycle. On-chip DLL generates 0˚, 90˚, 180˚, 270˚ four-phase clocks to produce the clock-centered output. Programmable impedance for on-chip input termination and output driver with precise linearity assure good I/O signal integrity at 1.75Gb/s/pin data rate. Programmable I/O skews allow user to compensate for board trace skews to improve the system performance.

Advanced system applications with high data throughput are always hungry for higher-BW SRAMs. Existing stand-alone high-speed SRAMs have a BW range of 50 to 80Gb/s [1, 2, 3]. In this work, a stand-alone 72Mb SRAM that can achieve 504Gb/s total bandwidth is presented.

Figure 34.3.1 shows the block diagram of the chip architecture and the features of this SRAM. The core is divided into 16 mats, each mat has two banks of 256k×9b, each bank has 32 sections. One section has 512 rows by 16 columns by 9 I/Os. During R/W operation both banks are accessed simultaneously to support DDR. For random_read and random_write within one clock cycle, one section will be selected for read during clock high, followed immediately by another (or the same) section for write during clock low.

The write drivers and sense amplifiers(SA) are located in the center of the mat to achieve balanced timings. The minimum cycle time is limited by core operation of read, write, and write-recovery all within one cycle. Self-timed R/W clocks and sensing control timing (shown in Fig 34.3.2) are designed with embedded core emulators to achieve minimum core cycle time.

A new SA is designed to meet the 3 tough requirements of this SRAM: wide data width(144_DDR), high frequency(>800MHz) and R/W multiplex. It employs a simple 2-stage dynamic SA for high speed yet low power, with a PMOS switch-pair as shown in Fig. 34.3.3. It consists of a pre-amp stage and a sense-latch stage with SA-out drivers. The switch-pair P1 and P2 perform multiple functions: buffer the line capacitances from SA1/SA1b, isolate pre-amp from write disturb, and precharge SA1/SA1b after read. These multiple functions are controlled by "psainb" signal. The cross-coupled N1 and N2 perform the first dynamic sense on SA1/SA1b activated by the current source NC1. The differential pair N3 and N4 work as dynamic sense signal repeater from SA1/SA1b to SA2/SA2b, by discharging SA2/SA2b at different rates to develop the differential voltage for the sense-latch stage. The main sense-latch stage is comprised of cross-coupled P3 and P4 pair(to Vdd), and N5 and N6 pair(to current sources NC2 and NC3). NC2 starts the sense operation of SA2/SA2b, followed by NC3 with additional current to speed up the SA2/SA2b splitting.

The dynamic SA has two key advantages: 1) Low power and high speed over a wide range of Vdd (0.6V to 1.5V), and 2) capable of multiplexing read with write in one cycle using the switch-pair P1 and P2.

In order to provide a clock-centered output, an on-chip DLL circuit is needed to generate 4-phase clocks: 0˚, 90˚, 180˚, and 270˚. 0˚ and 180˚ clocks are used to clock the output data while 90˚ and 270˚ clocks are used to generate echo clocks. Only one DLL circuit is used to save power and chip area. A conventional design uses 4 signal lines to distribute the 4-phase clocks across the chip to every I/O. However, process variation and device mismatch could create extra timing skews among the 4 signal lines. In this design, a signal line with 4 phases is used to minimize timing skews as well as reduce power dissipation. Figure 34.3.4 shows the DLL circuit with 4-phase clock generator. The push-pull circuit compresses 4 separate phase clocks into a single line with 4-phase. The 0˚ signal triggers P1 to generate the first rising edge (r1). The 90˚ signal triggers N1 to generate first falling edge (f1). Similarly, 180˚ and 270˚ signals generate second rising edge (r2) and second falling edge (f2).

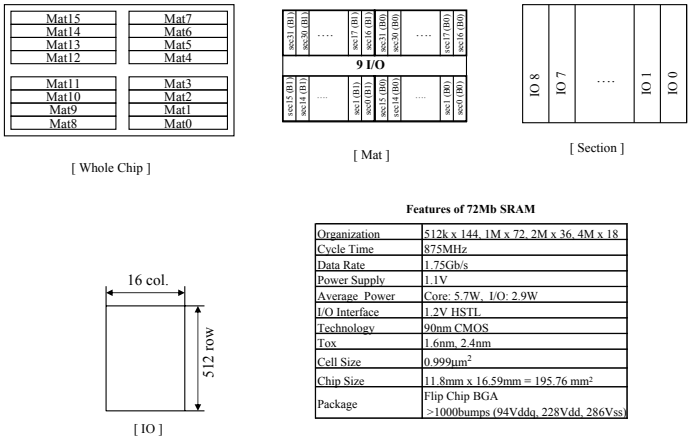
For a high-speed interface, precise input termination (ZT) and output driver impedance (ZQ) with good linearity is key to signal integrity. There are 4 critical factors to obtain good ZQ/ZT performance: 1) Linearity over I/O voltage range, 2) linearity over CMOS process variations, 3) low input capacitance (Cin) with robust ESD, and 4) glitch-less update. Solutions implemented in this design are:
1) Due to non-linear I-V curves of the MOS device, a combination of 90nm PMOS, NMOS and poly resistors are carefully modeled, simulated, and optimized in the design to achieve good linearity over the I/O voltage range. 2) Instead of $V_{ddq}/2$, this chip calibrates ZQ/ZT at V_{ddq}*80% and V_{ddq}*20%. There are two advantages of dual 80%/20% calibrations: First, for a terminated I/O interface, the Hi_Lo levels of I/O swing between roughly 80% and 20% of V_{ddq}. Dual calibrations at 80%/20% result in the precise control of VOH/VOL levels that are essential for the eye window. Secondly, PMOS and NMOS can be calibrated separately. Since at V_{ddq}*80%, pull-up NMOS is OFF while at V_{ddq}*20% pull-down PMOS is OFF. Hence, good linearity across PVT can be achieved. Figure 34.3.5 shows the linearity comparison of 80%/20% versus 50% calibration. 3) To achieve low Cin, it is critical to reduce CMOS driver sizes and its junction capacitance. Adding a series Poly resistor (Rpoly) to each 90nm CMOS driver reduces the junction area needed for ESD protection. Furthermore, an ESD device with low Cin is added in parallel to Rpoly to obtain robust ESD protection. 4) For glitch-less ZQ, data-dependent update of ZQ is used. ZT update is done by using thermal stacker code rather than binary code to avoid all bits switch.

For 144b data width, it is a big challenge for a board designer to match the trace delays of the data lines. As frequency gets higher, the timing budget of board trace skews has become a bigger portion of cycle time. This SRAM chip has programmable I/O deskews designed into every I/O with 16 programmable steps capable of 160ps de-skew per I/O, controlled by JTAG. Board designers can use this function to compensate for trace delay skews to improve the system performance.

Figure 34.3.6 shows the chip micrograph of the 72Mb SRAM. Figure 34.3.7 shows the data output waveforms with 1.75Gb/s/pin at 875MHz operating frequency and the shmoo plot. The waveform includes the data output as well as the echo clock.

Acknowledgements:
The authors like to thank G. Kao, J. Chang, K. Kohno, N. Urakawa, Y. Kobayashi and R. Veltman for their design support. The authors are also grateful to J. Kase, K. Matsuzono, L. Choi, A. Abusaidi and G. Merchant for their test support, and K.Fujita for his device support.

References:
[1] U. Cho, et al., "A 1.2V 1.5Gb/s 72Mb DDR3 SRAM," *ISSCC Dig. Tech. Papers*, pp. 300-301, Feb., 2003.
[2] H. Pilo, et al., "A 0.9ns Random Cycle 36Mb Network SRAM with 33mW Standby Power," *Symp. VLSI Circuits*, pp. 284-287, June, 2004.
[3] H. Pilo, et al., "An 833MHz 1.5W 18Mb CMOS SRAM with



Features of 72Mb SRAM

Organization	512k x 144, 1M x 72, 2M x 36, 4M x 18
Cycle Time	875MHz
Data Rate	1.75Gb/s
Power Supply	1.1V
Average Power	Core: 5.7W, I/O: 2.9W
I/O Interface	1.2V HSTL
Technology	90nm CMOS
Tox	1.6nm, 2.4nm
Cell Size	0.999μm^2
Chip Size	11.8mm x 16.59mm = 195.76 mm^2
Package	Flip Chip BGA >1000bumps (94Vddq, 228Vdd, 286Vss)

Figure 34.3.1: Chip architecture and features.

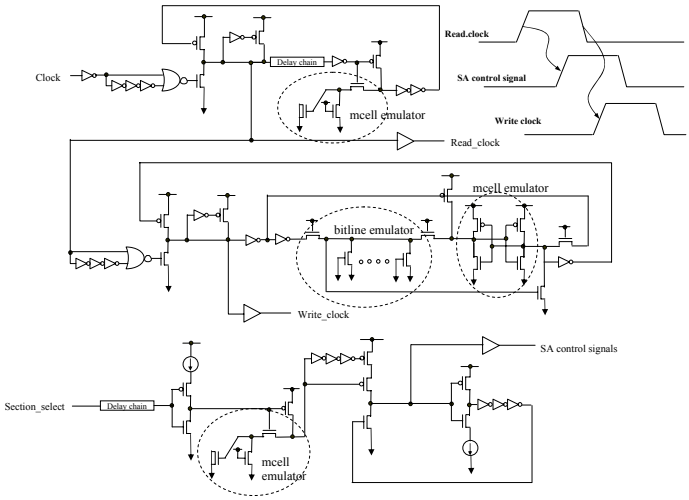


Figure 34.3.2: R/W clocks and sense control with core emulators.

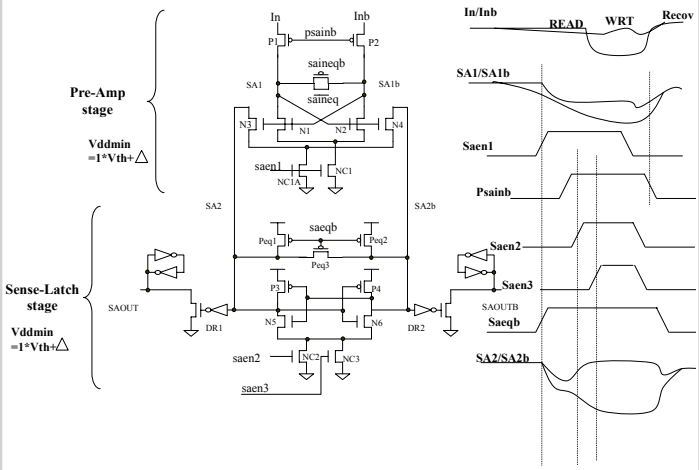


Figure 34.3.3: 2-Stage dynamic sense amp for 72Mb SRAM.

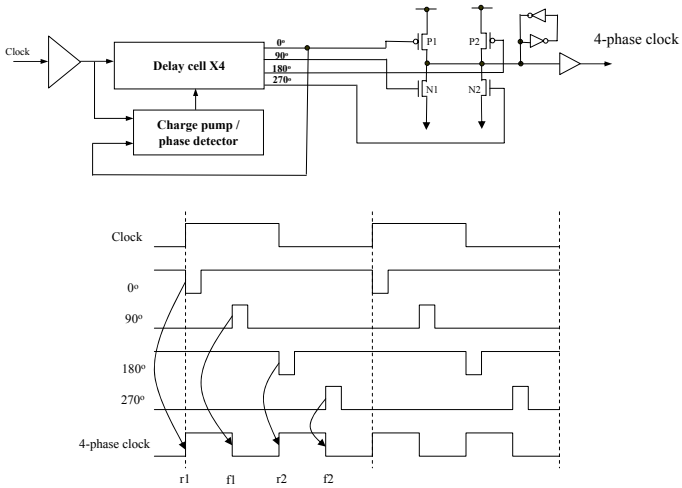


Figure 34.3.4: DLL with 4-phase clock generator.

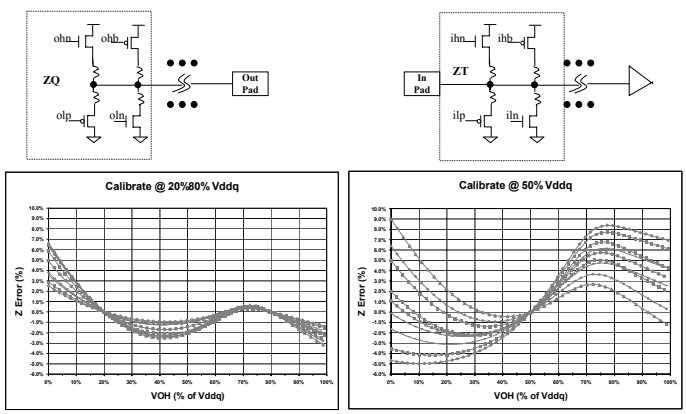


Figure 34.3.5: ZQ/ZT linearity comparison of 80%/20% versus 50% calibration.

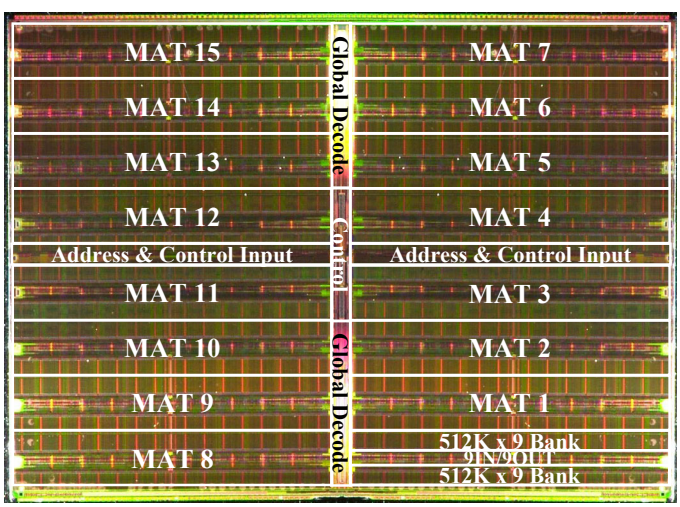
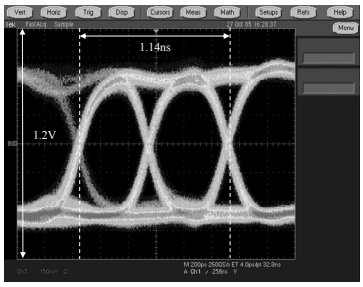


Figure 34.3.6: Chip micrograph of 72Mb SRAM with 144x2 I/O.

Continued on Page 678




```
DESKEW SET OK        ,00,00,02.0A, TKQ Begin  1.000NS  TKQ END  3.281NS  TKQ STEP  31.25PS
          0.00    0.25    0.50    0.75    1.00    1.25    1.50    1.75    2.00    2.25 [nS]
Vdd     |+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
1.300V  0000.111111111111111.000000000000000.111111111111111.000000000000000
1.280V  0000..111111111111111.000000000000000.111111111111111.000000000000000
1.260V  00000..111111111111111.000000000000000.111111111111111000000000000000
1.240V  000000.111111111111111.000000000000000.111111111111111.000000000000000
1.220V  000000..111111111111111.000000000000000.111111111111111.000000000000000
1.200V  00000000.111111111111111..000000000000000.111111111111111.000000000000000
1.180V  00000000..111111111111111..000000000000000.111111111111111.000000000000000
1.160V  00000000..111111111111111.000000000000000111111111111111.000000000000000
1.140V  000000000..111111111111111.000000000000000.111111111111111000000000000000
1.120V  000000000000..111111111111111.000000000000000111111111111111.000000000000
1.100V  000000000000..111111111111111..000000000000000.111111111111111000000000000
1.080V  000000000000...111111111111111..000000000000000111111111111111000000000000
1.060V  000000000000000..111111111111111...000000000000000.111111111111111.000000
1.040V  .000000000000000..111111111111111...000000000000000.111111111111111000000
1.020V  11.000000000000000..111111111111111..000000000000000.111111111111111100000
1.000V  1111.000000000000000..111111111111111..000000000000000.111111111111111
Vdd     |+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
          0.00    0.25    0.50    0.75    1.00    1.25    1.50    1.75    2.00    2.25 [nS]
```

Figure 34.3.7: Eye diagram and Vdd_tKQShmoo.

